April 2022

Methodology & Accuracy Summary

10m Global Land Use Land Cover Maps

Leaders in governments, NGOs, finance and industry need trustworthy, actionable information about the changing world to understand opportunities, identify threats, and measure the impacts of actions. The IO Land Use & Land Cover (LULC) annual maps, covering 2017 through 2021, meet these needs with timely, accurate maps at previously unobtainable scale. We use AI-powered algorithms to create 9-class global maps at 10m resolution and with over 75% accuracy.

Overview

The Impact Observatory LULC model leverages deep learning to infer a LULC class for each pixel in a Sentinel-2 image. Each annual LULC map is produced by aggregating a year's worth of inferences over the entire globe, processing approximately 2 million Sentinel-2 images to create each annual map. Maps for 2017 through 2021 have been published under a CC BY 4.0 license on both [Esri Living Atlas] and [Microsoft Planetary Computer] (as of April 2022 release).

Methodology

Our LULC classification algorithm is trained on a dataset of over 5 billion human-labeled pixels, based on 24,000 5 km x 5 km training sites that were distributed across all 14 major biomes. Expert labelers familiar with each biome labeled a subset of the data that was then used to train a non-expert group of annotators.

As described in our IEEE IGARSS paper[1], we employ a fully convolutional neural network with a UNet architecture to classify each individual image with a total of nine class outputs including water, trees, flooded vegetation, scrub, built area, rangeland, bare, snow / ice, clouds, and no data.

To create an annual map, we generate class inferences for all available Sentinel-2 imagery over each Sentinel-2 tile with our trained LULC model. This involves processing approximately 95 images for every location on the planet for each year—amounting to more than 2 million Sentinel-2 scenes and 2 petabytes of satellite data per annual map. We then aggregate the classifications across a year's worth of observations into a single annual composite.

[1] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis and S. P. Brumby, "Global land use / land cover with Sentinel 2 and deep learning," 2021 *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 4704-4707

To validate the accuracy of the outputs of our machine learning model, we compare the model outputs to human-labeled classifications over specific tiles. Our validation set, which is entirely excluded from model training, includes 398 areas of interests that collectively form a stratified random sample of the world's major biomes. Each labeled validation image spans 510 x 510 pixels and is labeled by three different expert annotators. These form a "gold standard" for validating our model results.

For the first version of our LULC map, released in 2020, we assessed accuracy using a standard "strict consensus" methodology. With strict consensus, the pixels used for validation are those where all three experts labeled the pixel and agreed on the feature class. Strict consensus provides a set of high confidence pixel classifications and includes fewer than 30% of all possible validation pixels.

As an alternative methodology, we compared our outputs to a "majority consensus" on the validation set, where at least two experts agreed on the feature class for the pixel, or one expert had an opinion and the other two did not. Majority consensus labels include nearly 80% of all possible validation pixels by incorporating lower confidence validation labels, and also result in "tougher grading" because they represent areas where land cover is harder for humans, and therefore machines, to reliably identify the LULC class.

Results

Accuracy Assessment

We assessed each of the 2017-2021 annual maps against the validation tiles. and found that each of the maps is at least 91% accurate when compared to a "strict consensus" and at least 76% accurate when measured using "majority consensus." For the detailed results that follow, we focus on "majority consensus" as it is a more stringent measure of our maps' performance.

	Precision [%]	Recall [%]
water	87.2	90.1
trees	82.3	85.4
flooded veg	57.3	53.9
crops	90.2	72.1
built area	79.6	94.5
bare	72.1	38.2
rangeland	57.8	70.5

Figure 1. 2021 majority consensus results.

Precision and Recall

Precision and recall are key performance metrics used in machine learning to estimate the utility of a classifier. Precision (or user's accuracy) reports the fraction of predictions that were correct and recall (or producer's accuracy) reports the fraction of validation examples that were correctly classified. The precision and recall of each class is evaluated over the majority consensus validation set. The results for 2021 are summarized in Figure 1. These results are generally similar to those of the 2017-2020 maps.

Confusion Matrix

A confusion matrix graphically summarizes the occurrence of individual classes being correctly identified (the diagonal elements) and also the occurrence of individual classes being mistaken for another (the off-diagonal elements). The confusion matrices for the 2021 annual map when compared against 'majority consensus' validation data are depicted in Figure 2. Again, results for 2021 are generally similar to those for 2017-2020.

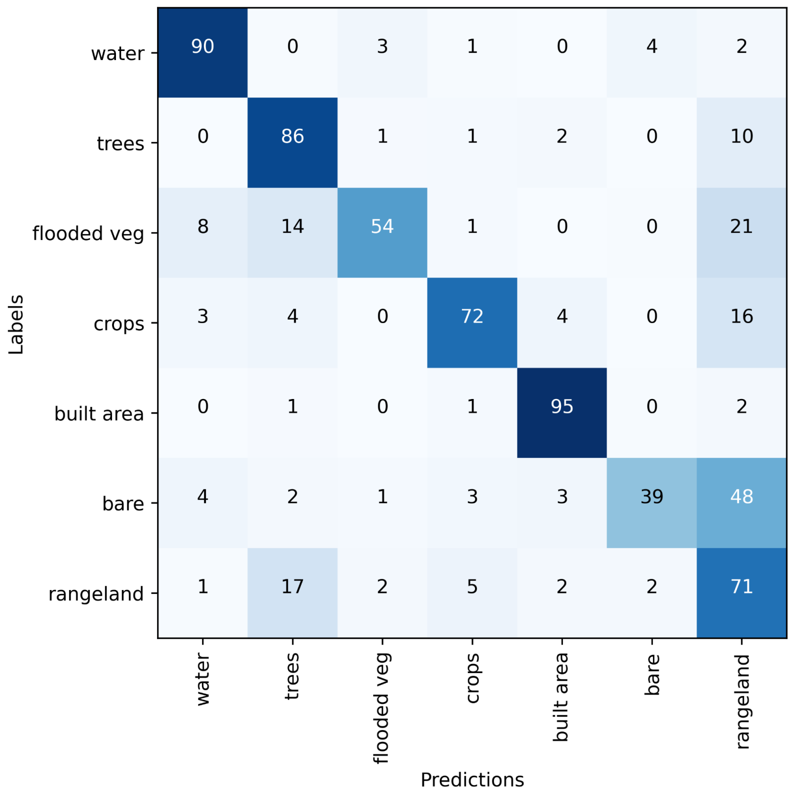

Figure 2. Confusion matrix from the 2021 majority consensus results.

For More Information

For more information on our annual time series of LULC maps, please see our website at https://www.impactobservatory.com/global_maps or contact us at hello@impactobservatory.com.